ABSTRACT OF THE DISCLOSURE

In this invention, when frame encoded data is to be generated on the basis of data obtained by separating image data and sound data contained in frame data of a motion image and hierarchically encoding both the data, frequency subbands of the same significance level in the hierarchically image encoded data and sound data are grouped, and frame encoded data is generated by arranging these groups in descending order of significance level. This makes it possible to appropriately give scalability to both the image data and sound data already hierarchically encoded, without decoding them, and generate encoded data containing both the data. Since encoded data of image data and sound data can be transmitted by grouping them in appropriate units, the receiving side can efficiently utilize the encoded data.